2013-02

Evidence-based programs in children's services: A critical appraisal

Axford, Nick

10.1016/j.childyouth.2012.10.017
Children and Youth Services Review
Elsevier BV

Evidence-based programs in children's services: a critical appraisal

Nick Axford and Louise Morpeth

The final version of this article is available here:
https://ac.els-cdn.com/S0190740912003994/1-s2.0-S0190740912003994-main.pdf?_tid=9f0c937b-bbed-4506-941e-09b6be4494c8&acdnat=1521667200_0919c61a9cb040d7308b9eac70013373

Citation: Axford, N. and Morpeth, L. (2013) Evidence-based programs in children's services: a critical appraisal. *Children and Youth Services Review* 35, 268-277. http://dx.doi.org/10.1016/j.childyouth.2012.10.017

Abstract

Evidence-based programs (EBPs) are increasingly being implemented in children's services agencies in developed countries. However, this trend is meeting resistance from some researchers, policy makers and practitioners. In this article we appraise the main critiques, focusing on scientific, ideological, cultural, organizational and professional arguments. We contend that some of the resistance stems from misconceptions or an oversimplification of issues, while others represent valid concerns that need to be addressed by proponents of EBPs. We set out implications for the development and evaluation of programs and how they are introduced into service systems, and conclude with broader recommendations for children's services.

Key words: Evidence-based practice; evidence-based program; children's services; prevention; early intervention; implementation, evaluation

Highlights:

- The scientific studies behind evidence-based programs must be reported clearly, as recommended in the CONSORT and TREND statements
- Evidence-based programs can further the cause of social justice
- Programs can translate across cultures but more empirical research is needed to know which ones and why
- Programs need to be 'system-ready' and systems need to be 'program-ready'
- Evidence-based programs can help professionals improve and enjoy their practice
- An international standard of evidence would give the term 'evidence-based' more credibility

1. Introduction

Recent years have seen the rise of the concept of the 'evidence-based program' (EBP) in children's services. While few people argue that services should not be based on evidence, EBPs have not been widely welcomed or adopted. There has been considerable resistance from practitioners, policy makers and academics.

In this article we aim to summarise the main critiques clearly and fairly and then appraise them. They are categorized as scientific, ideological, cultural, organizational, and professional. They are drawn from the literature and our experience over nearly a decade of adapting, implementing and evaluating EBPs in the UK, Ireland and the US. Our work has is predominantly been in social care, early years, education and youth justice and this is reflected in the examples we draw on. We endeavour to expose what we believe to be myths about EBPs but also to identify valid concerns that proponents of EBPs need to address. In doing so we set out implications for the design and evaluation of EBPs, the way in which they are introduced into service systems and the development of children's services more generally. Our comments are aimed at EBP supporters and skeptics alike, including fellow researchers, policy makers, practitioners and commissioners.

We should be clear about definitions. By 'program' we mean a discrete, organized package of practices, spelled out in guidance – sometimes called a manual or protocol – that explains what should be delivered to whom, when, where and how. A program is 'evidence-based' when it has been evaluated robustly, typically by randomized

controlled trial[1] (RCT) or quasi-experimental design[2] (QED), and found unequivocally to have a positive effect on one or more relevant child outcomes. EBPs sit within the broader evidence-based practice movement, which emphasizes using the best evidence available when intervening with children and families (Sheldon & Macdonald, 2009).

We should also be upfront about our stance. We support the greater use of EBPs in children's services. These programs have considerable potential to improve child well-being and are currently greatly underused, particularly given the relative dearth of evidence of impact for most 'services as usual'. But, as will be seen, they also have limitations and are certainly no panacea. Moreover, not all programs need to improve children's health and development demonstrably – other important objectives include upholding children's rights, enhancing their quality of life, and furthering inclusion (Axford, 2008) – and by no means all services should be programs.

In what follows we look at each of the five critiques in turn, first raising the issues and then offering a response. We hope that engagement with common critiques of EBPs, even if necessarily brief, will stimulate further constructive discussion on the subject, as well as sponsoring the greater take-up of existing programs where appropriate and a new generation of innovative programs proven to benefit children.

2. Scientific critiques

[1] A randomised controlled trial is a method for evaluating the impact of an intervention on selected outcomes. It involves randomly allocating potential beneficiaries of an intervention to a program or treatment group (who receive the intervention) or to a control group (who do not). Outcomes for the two groups are then compared.

[2] A quasi-experimental design is a non-randomized evaluation method in which participants referred to the intervention are compared with a group of matched participants who do not receive the intervention.

There are several philosophical and methodological concerns about the scientific methods that underpin the design of programs and the assessment that they 'work'. An argument that potentially undermines the EBP concept is that the philosophical basis for determining that a program is 'evidence-based' is fundamentally flawed. Relying solely on empiricism – that is, the notion that knowledge comes only or primarily from measurable experience – is considered to underestimate the value of other ways of knowing (Webb, 2001). RCTs meanwhile, which are commonly regarded as the 'gold standard' evaluation method in the hierarchy of evidence and therefore regularly used as *the* marker of an EBP, are deemed to be inadequate for understanding the complex issues that surround intervention effectiveness (Gray et al., 2009; Stewart-Brown et al., 2011). It is naïve, it is claimed, to believe that programs simply work or not (i.e. that they are 'evidence-based' or not); rather, the question is whom they work for, why and in what contexts (Pawson & Tilley, 1997).

Allied to this is the argument that much existing provision is not necessarily ineffective just because it has not been tested by RCT. It may not be amenable to experimental evaluation, for example because it is not manualized, and much practice has been shown by longitudinal studies to be effective (Thoburn, 2010). Similarly, there are many 'home-grown' programs, mostly sponsored by the voluntary and community sectors. These are often popular with users, who report being helped by them, and a growing proportion of such interventions have been held up as examples of good practice, or 'validated' (though not subjected to a formal measurement of impact).[3] There is a danger, it is contended, of replacing effective 'business as usual'

[3] For example, www.c4eo.org.uk/themes/general/localpracticeexamples.aspx?themeid=10.

with EBPs that may be unsuitable in a context that is different from where they were tested and found effective.

Our view is that RCTs are by no means the only method that should be applied: it depends on the question being asked (Bullock et al., 1992; Little et al., 2005). Studies looking at, for instance, what it feels like to receive a certain service or why clients drop out of a program require surveys or qualitative methods, not experiments. But RCTs are the best means available of identifying impact because they control for extraneous factors (Torgerson & Torgerson, 2008). They are not always ethical or practical, and there are other methods of impact evaluation (Rossi et al., 2004). QEDs can be an affordable or practical alternative, and there are techniques like propensity score matching for improving the comparability of groups in such studies. However, such methods have important limitations, notably their relative complexity, their need for large samples and their failure to account for unmeasured differences between groups (Shadish, 2012).

What of the argument that much effective provision already exists and is in danger of being displaced unnecessarily by new programs? First, pre-post evaluations or qualitative studies showing that users found a service helped them have limited validity because benefits for recipients cannot confidently be attributed to the program (Ritter, 2012). Yet this is the highest level to which most 'home grown' innovations in the UK have been evaluated. The result is that claims regarding the effectiveness of current practices are often over-stated. It has been argued persuasively that if the public, as taxpayers, knew the extent of this they would demand higher quality evaluations (Farrington, 2003). The solution, then, is not to lower the standard of what

constitutes ‘evidence-based’, but nor is it to subject every new program to an RCT. Rather, the most promising innovations should be strengthened and tested with a level of rigor appropriate to their stage of gestation as part of a logical process of program development (Little, 2012). Those early in their gestation might warrant a pre-post or even a small comparison group study, with progress to larger randomized trials conditional on promising results.

Second, until now RCTs have not had privileged status amongst policy makers (Ritter, 2012). As a result, EBPs are likely to constitute only a small proportion of total spending on children’s services for the foreseeable future – not so much taking over the world as scratching at its surface (Little & Axford, 2012). It is highly probable that much current social care, education, health and youth justice provision – effective or not, evaluated by RCT or not – will survive for many years yet.

Scientists who accept the value of empiricism and experimental evaluation to determine whether or not child outcomes are affected by an intervention nevertheless point out that trial results, even from rigorous trials, can mislead. This causes some programs to be badged erroneously as ‘evidence-based’. One criticism is that in trials finding no overall effect investigators often conduct multiple sub-group analyses and claim a program is effective because it benefited a discrete group within the sample – perhaps females of a specified ethnic group, or those who received a high-fidelity version of the program (Gorman, 2002a, 2005a). A further problem with many EBPs, say critics, is that they are labelled ‘effective’ on the basis of an impact on only a small proportion of the total number of outcomes measured (Gorman, 2002b). In other words, there is no preponderance of evidence in their favor. It is suggested that

researchers also sometimes reduce attrition[4] rates artificially, for instance by measuring them relative not to the point of randomization but to a later point, by which time some cases have already dropped out. This is a problem because high attrition creates possible selection bias – participants lost to a study may have different characteristics to those that stay in the study – and therefore undermines confidence in any observed effect of the program (Gorman, 2002a, 2005a). Another error is to use one-tailed statistical significance tests, which doubles the likelihood of finding a statistically significant effect, even though their use should be limited to instances in which previous research has led one to expect only positive, not negative, effects (Gorman, 2003). In short, there often appears to be an inbuilt bias to find a positive effect (Holder, 2009).

Many of these criticisms are aired in a fascinating exchange between Dennis Gorman and the developers of two widely respected EBPs – *Raising Healthy Children* and *Life Skills Training* (Gorman, 2005a/b; Hawkins & Catalano, 2005; Botvin & Griffin, 2005). It is clear from this that the devil is often in the detail of program evaluations, so caution is needed when making claims about impact, especially if effects are patchy – there is inconsistency between studies or across sub-groups, outcomes, follow-up period, level of fidelity and so on. Often, judging what the preponderance of evidence is saying is just that – a *judgment*. Scientists disagree on the findings that emerge from RCTs and on how much and what type of evidence is needed to certify something as 'evidence-based'. This reinforces the need for better reporting of trials and for clear standards of evidence at national and ideally international levels. Care is

[4] Attrition is the loss of participants to a research study. In the case of an experimental evaluation this could be from the comparison group or intervention group and means that outcome data is not available for that participant. High levels of attrition reduce confidence in a study's findings.

also needed to avoid the danger of seeking to *verify* rather than *test* hypotheses concerning possible effectiveness. In our view, the proper response to the points made is not to reject EBPs as a category but to improve the standard of evaluations, apply stringent standards of evidence in clearinghouses[5] of EBPs, strengthen the rigor of reporting of trials and register *all* trials on a public database.[6]

A related scientific critique comes from the increasing body of evidence showing that evaluations in which the program developer is involved are more likely to show positive findings than those conducted by independent researchers (Eisner, 2009a). For example, there have been numerous positive evaluations of the *Triple P* Level 4 group parenting program but to our knowledge only four RCTs have been undertaken independent of the program originator (Gallart & Matthey, 2005; Hahlweg et al., 2010; Malti et al., 2011; Eisner et al., 2012; Little et al., 2012). Viewed together these show equivocal evidence of impact on child development. A generous interpretation is that fidelity and quality are better when designers oversee implementation, but the possibility of systematic bias in the design, analysis and interpretation of such evaluations cannot be dismissed.

In response, it should be noted that alternative explanations have been offered. One is that there could be a 'prosecutor bias' whereby independent evaluators are corrupted by the desire to disprove the value of respected programs (Sherman & Strang, 2009). Another is that, for various reasons, programs may be evaluated in the real world

[5] There are several online clearinghouses of EBPs. Examples include Blueprints for Prevention – www.colorado.edu/cspv/blueprints/ and the Best Evidence Encyclopaedia – www.bestevidence.org/.
[6] The International Standard Randomised Controlled Trial Number Register (http://isrctn.org/) is a database of trials enabling users to track studies from initial design to publication of results. Trials registered must supply information for various data fields, including target group participants and primary and secondary outcome measures.

prematurely, so that poor results reflect the need for further development or better implementation (Olds, 2009). The pattern is not universal either, as demonstrated by the 'black swan' case of *negative* developer bias; a review of 12 restorative justice trials found that those analyzed by independent evaluators showed more benefits than those analyzed by developers (Sherman & Strang, 2009). Rather than dismiss all developer-led evaluations as biased, there should be greater transparency about potential financial and ideological conflicts of interest (Eisner, 2009b). This would help to minimize potential biases related to developer involvement, or least allow them to be better assessed. Further, some clearinghouses of EBPs specify that in order to qualify a program must be found effective in an independent evaluation.

It is also often unclear what actually happened in a trial, making it difficult to interpret findings. Analyses of RCTs in crime prevention, for example, have found that randomization method, intent-to-treat analysis, differential attrition and possible adverse effects – all of which are important for judging the strength of evidence of effect – are generally under-reported (Perry et al., 2010; Gill, 2011). We agree that transparency is essential when reporting evaluations. The CONSORT[7] and TREND[8] guidelines, essentially checklists to ensure that randomized and non-randomized trials respectively are reported in a standard way to aid their critical appraisal and interpretation, should therefore be widely adopted by evaluators and publishers.

The other main scientific critique concerns the risk factor paradigm that underpins most EBPs – the quantitative pursuit of the causes and predictors of developmental problems (eg. Case, 2006, 2007). The research used to identify risk factors is deemed

[7] www.consort-statement.org/
[8] www.cdc.gov/trendstatement/

to be self-fulfilling in that it uses a prescribed menu of factors derived from research with a narrow psychosocial focus. It tends to neglect structural and political influences, such as poverty and inequality, in favor of family, school, peer and neighbourhood factors. There is also a reliance on 'objective' measures of 'adult-prescribed' risk, with insufficient qualitative investigation of how children perceive them.

This critique is about how the risk factor paradigm is used rather than about its validity *per se*. First, the paradigm does not necessarily lead to *targeted* programs: many if not most universal EBPs are also informed by the risk factor paradigm. *Guiding Good Choices*, for example, a program for parents of 8-14 year-olds that aims to reduce substance use initiation (Haggerty et al., 1999), is informed by a comprehensive review of risk and protective factors (Catalano & Hawkins, 1996). Second, while there is compelling evidence in western developed countries for many such factors, the need to understand local specificity – for instance, through community surveys – is readily conceded, as is the need to elicit the meanings that young people attach to risks (Little et al., 2004).

On the question of the *validity* of the risk paradigm, proponents would say that it works. In youth justice, for example, effective programs specifically address key risk factors: they focus on building more effective family functioning, disengaging youth from deviant peer networks and enhancing school performance. Ineffective programs largely ignore these risk factors, or make them worse – for example, by putting antisocial youth together for extended periods of time (Henggeler & Schoenwald,

2011). This is not to say that structural factors like poverty or inequality are unimportant or should be ignored, as the next section shows.

3. Ideological critiques

Ideological critiques concern the way in which EBPs purportedly embody and further right-wing values, both failing to address structural causes of child well-being problems and, in some cases, pursuing ends that are illiberal and regressive. We start with the view that such programs are bound up in a neoliberal approach to children's services that emphasises the role of the market. This places EBPs within a broader set of policy reforms that are deemed to be preoccupied with preventing anti-social behavior and saving public money or even generating private profit (Garrett, 2009). An associated critique concerns a perceived obsession with controlling and monitoring frontline practice, holding services accountable and ensuring value for money (Chard & Ayre, 2010). Further, many EBPs emanate from the US, and there is a worry that the welfare state in other countries is being *re-made* in the image of US-style 'service modernisation' efforts – for example so that services are more businesslike and outcomes-orientated.

Several points may be made in response. First, EBPs are by no means all concerned with reducing anti-social behavior. For example, there are programs for literacy, job skills, mental health and many other developmental outcomes. Second, good stewardship of resources involves diverting expenditure from less effective to more effective intervention and, when economically sensible, shifting resources from intervention to prevention. When public finances are being squeezed there is even more pressure to know the cost and benefit of interventions. Many EBPs can help

reduce unnecessary future expenditure on remedial provision (Aos et al., 2011). Third, the concern about surveillance assumes that EBPs necessarily require much additional data collection. This is not the case, although the collection and continuous feedback of real-time data (eg. Bickman et al., 2011; Cash et al., 2012) are arguably the best way to fulfil practitioners' desire and a manager's need to know that what they do makes a difference (cf. Stalker et al., 2007).

Further, the concern about a focus on outcomes is, in our view, misplaced. It tends to be associated with an emphasis on performance indicators, yet used properly it should divert attention away from outputs and towards improving child well-being. A concern with efficiency, effectiveness, and clear goals is good management, *not* 'managerialism'. Moreover, an outcome focus advances user empowerment and civil liberties. Work to help children and parents resolve their difficulties should be well targeted and not wasted, and intrusion into family life must be properly justified. Professionals work best when they understand their purpose and how their activities contribute, while clients feel more confident about an intervention and hopeful of its success if they understand why things are being done to them (Thompson, 2008).

Another ideological critique is that the causes of problems afflicting children and families in many developed countries are primarily structural – economic inequality and political disenfranchisement – yet EBPs do nothing about these. Instead, they tend to individualize problems: cognitive behavioral therapy, for example, treats the symptom, namely depression, not the cause, which could be overcrowding, poverty or isolation, and most parenting programs do or say little about discrimination, inequality or social deprivation (Forrester, 2010).

Cross-national analyses reinforce this argument. Countries with the worst child well-being are very unequal economically, whereas those with the best child well-being are more equal (UNICEF, 2007; Wilkinson & Pickett, 2009). The former, in particular the US, make greater use of EBPs, which tend to be reactive to identified risks and proliferate in the context of limited (and often poor) universal services. Countries topping international league tables, by contrast, adopt more redistributive policies and provide universal welfare and benefits systems as a right (Rowlands, 2010). They foster a social context in which, compared with the UK, children spend more time with family and friends, engage more in active leisure pursuits and, along with their parents, are less materialistic (Ipsos MORI, 2011). Mostly concentrated in northern Europe, high-achieving countries also invest in professionalizing the children's services workforce, notably through social pedagogues who are trained in child development and work therapeutically with children and families in many settings (Petrie et al., 2006). Similarly, violent societies tend to have high inequality, unlike more peaceful societies (Eisner & Nivette, 2012).

We agree that the evidence that higher societal inequalities are associated with poorer outcomes for children is incontrovertible (Wilkinson & Pickett, 2009). Without wider structural reforms EBPs will always be swimming against the tide in their efforts to improve child well-being (Rowlands, 2010). It is also true, as critics point out, that most EBPs originate in the US. But this is not evidence that EBPs are responsible for that country's social ills. The emphasis in the US on designing and testing interventions derives in part from philanthropy, with its concern for evidence that money is well spent, providing high-quality interventions and sponsoring innovation

in the context of poor or non-existent public services. Public concern about high rates of violence, drug misuse and juvenile detention has also stimulated investment in establishing and disseminating 'what works' in prevention and early intervention (Elliot, 2010).

EBPs tend not to address structural factors, but this does not justify abandoning them. It simply highlights one of their limitations (one that arguably applies to most frontline services). This said, EBPs can and arguably do contribute to greater social justice. First, both universal and targeted programs can narrow the gap between children in greatest need and their peers. Often children with the most serious problems make the largest gains (eg. Tolan et al., 2004). Second, the links between poverty and outcomes are complex and mediated by various other factors, such as parenting behavior. The task of making significant inroads into or even eradicating poverty or extreme inequality is likely to take a generation at least and requires policy intervention. By targeting mediators of poverty – for instance, by improving parenting skills – EBPs can make a difference to child well-being *now*. Third, EBPs *can* intervene at a structural level. The *FAST (Families and Schools Together)* program, for example, helps to build social capital and address discrimination insofar as implementers must be broadly representative of the community (Kratochwill et al., 2004).

A further ideological criticism of some EBPs is that they address problems that do not exist, or at least seek questionable ends, and are therefore superfluous and even dangerous. For example, school-based social and emotional learning curricula such as *PATHS (Promoting Alternative Thinking Strategies)* are deemed to make children see

themselves and others as fragile and vulnerable, thereby promoting dependence on state agencies for emotional support (Ecclestone & Hayes, 2009). There are similar debates elsewhere in prevention science, for example over what constitutes 'problem drinking' (Forrester, 2010).

One response is that levels of emotional and behavioural problems among children in the UK are significant and, until recently, rising (Maughan et al., 2008; Collishaw, 2012), and improved social and emotional well-being contributes to more successful learning (Durlak et al., 2011). There have been rejoinders to such points (Ecclestone & Hayes, 2009) and the exchange of views will undoubtedly continue, as healthy discussions should. It is incumbent on developers and promoters of EBPs to take a critical view of the definition of issues being addressed (Forrester, 2010) and to be open to the possibility that some programs are harmful – *Scared Straight* being a notorious example (Petrosino et al., 2003). It is unethical for children to receive such programs – and of course, they are not 'evidence-based' in the sense we described earlier – yet many still do.

4. Cultural critiques

A concern often voiced about EBPs is that they do not translate across cultures. Specifically, a program tested and found effective in one country may not work in another country, and a program that benefits one ethnic group will not necessarily benefit others.

A variety of evidence is offered to support this view. Some is essentially anecdotal, and concerns the 'face validity' of program materials. Practitioners review program

manuals and DVDs and conclude that the language, images, stories, voices and faces would alienate their service users. Other evidence concerns problems with implementation, such as low fidelity or difficulties engaging parents (Kumpfer et al., 2012). There are also cases where programs have been evaluated in a country beyond where they originated but with mixed or disappointing outcomes. For instance, when *Multisystemic Therapy*, which was developed in the US, was implemented in Sweden it appeared to benefit participants but no more so than services as usual (Sundell et al., 2008). As a result, it failed to generate the anticipated financial benefits (Olsson, 2010). By contrast, an RCT in the UK that compared MST with business as usual *did* find that MST outperformed services as usual on some outcomes (Butler et al., 2011).

Many of the best-known EBPs have been developed in the US. The contrast between the US and other countries is often used as a reason to resist US programs on the grounds that they will not suit a different context. In Europe, states with social-democratic or Catholic welfare regimes tend to be instinctively skeptical about interventions developed in the US (Grietens, 2010).

Several reasons are given for the apparent lack of fit of EBPs in new countries or with different ethnic groups. The first is that the nature of the problems that are targeted and their etiology are different. For example, it is claimed that families in each country have unique needs, and the risk and protective factors that contribute to them are culture-specific, not universal (Kumpfer et al., 2012). Similarly, programs that promote 'western' models of parenting, such as non-directive play and non-corporal punishment, may not be seen as desirable or relevant in some cultures. Second, there are aspects of program designs that jar with indigenous forms of policy and practice.

These might concern underlying assumptions, presentation, duration, organization, language, accent, expectations of recipients, and so on. For instance, the idea of parent training programs does not sit easily in France, where parents are considered the experts in raising their children (Milova & Sohre, 2011). Third, indigenous service systems may not be set up to house certain programs. The type or level of practitioners' training and their pedagogic style, for example, might make it difficult and even impossible to deliver a given EBP well.

From our perspective these are all valid concerns. However, transportability is essentially an empirical issue, and empirical evidence suggests that some nuance is needed. First, some programs appear to be more transportable than others. The consistent positive findings for *Incredible Years* in varied contexts and in many countries are striking (eg. Webster-Stratton et al., 1988; Hutchings et al., 2007; Larsson et al., 2009; Scott et al., 2010; Little et al., 2012; McGilloway et al., 2012). Rather than write-off all programs developed abroad, it makes sense to learn what makes *Incredible Years* more transportable than some other EBPs. For example, a study in London showed that it successfully overcame cultural and ethnic barriers to engagement. Minority families engaged as well as Caucasian ones, were just as satisfied, and showed equal changes in their parenting (Scott et al., 2010). The researchers attributed this to the program's sensitivity in addressing cultural issues. There are other examples of successful implementation across contexts, including Parent Management Training in Norway (Ogden & Hagen, 2008) and Nurse Family Partnership in England (results of the trial of the latter have not yet been published). A study involving interviews with several adopters of US-developed programs in

Europe showed that their implementation experiences and outcomes were generally positive (Ferrer-Wreder et al., 2004).

Second, it is simplistic to say that the epidemiology of developmental problems is culture-specific. Generally, EBPs are based on child development processes that apply across cultures and national boundaries (Little & Maughan, 2010; Maughan & Little, 2010). There is some variation, particularly when human agency and meaning are accounted for, and this is larger for some problems, such as alcohol misuse, than others – say conduct disorder. But the commonalities outweigh the differences, at least in the developed world (Chen & Eisenberg, 2012; Eisner & Nivette, 2012). For example, cross-cultural studies show that warm but firm parenting is associated with better child outcomes irrespective of culture, and that parenting programs developed in the West may be universally applicable and effective (Scott et al., 2010).

Third, the differences in impact between trials in exporting and importing countries respectively may be explained by other factors besides culture. Studies in importing countries are less likely to involve program developers, who appear to influence impact. The level and quality of 'services as usual' affects the added value of the program: if they are extensive and good, the effect size will likely be deflated (Olsson, 2010). These factors reinforce the need for clearer reporting in trials, for example rigorously detailing what participants in control conditions actually receive.

Clearly, though, the process of transporting a program from one country to another, or applying it to different ethnic groups, requires careful thought. Surprisingly, methods for adapting programs culturally have not been tested (Kumpfer et al., 2012), even

though the principle of starting to develop ways of working in one place, then looking to see if it works elsewhere and critically examining the impact of context, seems sensible (cf. Pawson & Tilley, 1997). But there is reasonable consensus that programs can be adapted in a planned fashion, and that this is best done in collaboration with practitioners, intended users and the program originator (Bumbarger & Perkins, 2008). Changes should be made gradually and tested continuously (Fraser & Galinsky, 2010). The challenge is to preserve the underlying logic model since eliminating or modifying core components will likely reduce program effectiveness. Instead, the focus should be on adjusting culture-relevant language, colloquialisms and examples, recognizing norms of accepted and undesirable behavior. These are more surface accommodations than deep adaptations, and, unless consumer groups need or want new content, it is generally accepted that it is the form of delivery – personnel, mode of communication, location – that changes (Misurell & Springer, 2011).

There are limits to what can be done, however. For example, research on adapted versions of *Strengthening Families*, a universal program to reduce anti-social behavior and alcohol and drug abuse amongst 10-14s, found that culturally-specific versions with various populations increased recruitment and retention of families by 40% on average (Kumpfer et al., 2012). However, they did not yield significantly improved outcomes. Further research is needed into how to achieve better outcomes and further justify the cost of cultural adaptation. There is also the practical problem that in multi-cultural, multi-ethnic samples the intervention cannot be tailored to address every group (Scott et al., 2010). There is much to be said for programs that

are generic but flexible enough for experienced facilitators to individually tailor to individual families' needs while maintaining core elements.

We offer one final response to the cultural critique of EBPs, which is the need to develop programs in Europe and elsewhere and promote them more widely. Several exist (e.g. Atria & Spiel, 2007; Faggiano et al., 2008; Salmivalli et al., 2011; Stemmler et al., 2013) but they need to become better known, which calls for clearinghouses of EBPs to be in different languages and aimed at diverse cultures (Soydan et al., 2010).

5. Organizational critiques

A critique gaining momentum maintains that most EBPs were developed and found to work in rarefied conditions, that they do not translate into real-world children's services systems, and that therefore alternative approaches are needed. To start with, critics contend that EBPs are rarely tested and found effective under 'real world' conditions (Thoburn, 2010). It is true that many RCT and QED evaluations, typically led by universities, have features that are not replicated in public systems. First, the program developer is often involved in the delivery and evaluation (Eisner, 2009a). Second, the research team, generally educated to doctorate level, may also be involved with delivery. This may make them differently qualified and motivated to those who would normally do so. Third, the programs are often tested on homogenous groups, thereby excluding users with co-morbid conditions or other complications – they 'work' but not with 'real' children. Fourth, the evaluations are typically well resourced, meaning that there is a good budget for training, materials and wages. Such

conditions are arguably unlike the day-to-day reality of children's services and limit the generalizability of results from studies to public systems.

We believe that these are all good reasons to be cautious about the applicability of EBPs to public systems. However, where programs have been tested and found effective across multiple settings in real world conditions – meaning that they are delivered by regular staff in orthodox service settings and without intensive support from people employed by the research team – the case against their wider take-up is weaker. This is particularly so when much routine provision for children has not been robustly evaluated. Any program that achieves this standard, such as those in the Blueprints clearinghouse, should be deemed not only 'effective' but also 'ready for dissemination' in communities and public service systems (Axford et al. 2012a).

Some critics of EBPs accept that programs can make a demonstrable impact on children's lives but question the ability of service systems as currently configured to provide an appropriate home for programs. The evidence offered in support of this is that hardly any public system in the world has imported and implemented an EBP at scale – that is, with high market penetration (Little, 2010). For example, *Family Nurse Partnership* has probably had the greatest impact of all imported EBPs in the UK, yet its current penetration rate is estimated to be about 20%. This is a significant achievement but it is by no means scale.

One way of analyzing the barriers that systems present is to consider their 'readiness for implementation' or, put another way, to ask whether the EBPs are being planted in fertile ground (Fixsen et al., 2005; Bumbarger & Perkins, 2008). Typical challenges

include short-term funding, failure to invest in organizational infrastructure, a lack of fit between the goals of the program and the goals of the system, and weak processes for recruiting and retaining participants. These are only a few examples.

Little's (2010) challenge that public systems tend not to be well suited to EBPs and that EBPs are often not designed with systems in mind seems reasonable to us. Rather than reject the concept of EBPs altogether, though, program developers need to re-focus their design efforts to think about the challenge of scale from the outset. This includes providing empirical evidence for the elements of programs that are core and those that are amenable to local adaptation, and acting on lessons concerning the effective diffusion of innovations (Little, 2011).

Some prevention scientists, aware of the difficulties of fitting programs into systems, have argued for evidence-based *practice* rather than programs. Lipsey (2010) suggests looking at what can be learnt from meta-analysis to develop 'specific operating procedures'. Rather than select a single parenting program, for example, a meta-analysis would identify good practice from across numerous such programs. One review identified seven critical ‘features’ of parenting interventions to address conduct disorder (NICE, 2006). It recommended, for example, that the work be based on social learning theory, use group settings and include role-play and homework so that parents can practice what they have learnt. Whereas an EBP would usually prescribe exactly what is done, by whom, when and how, this approach relies on staff being able to work with more general guidance to shape practice.

A similar approach involves identifying individual strategies to influence behavior.

Embry and Biglan (2008) define 'evidence-based kernels' as 'fundamental units of behavioral influence that appear to underlie effective prevention and treatment for children, adults, and families' (p75). Examples include time-out, nasal breathing and verbal praise. This approach – also referred to as 'practice elements' (Mitchell, 2011) or 'common elements' (Barth et al., 2012) – allows practitioners discretion to combine practices in tailored packages according to children's presenting needs.

These strategies need not be in competition with EBPs but we would note that they are not unproblematic. Practitioners who are less able arguably need the more explicit guidance provided in carefully constructed programs. Moreover, in some subject areas there is a dearth of high-quality evidence, so attempts to distil critical practices may, paradoxically, end up relying on a few EBPs. For example, in a recent systematic review of parenting programs for conduct disorder nine of the 13 qualifying studies concerned *Incredible Years* (Furlong et al., 2012). Even so, in many areas the evidence does not yet provide a clear steer for public systems on the relative effectiveness of programs, practices and kernels and the associated costs. There must be scope within the vast majority of children's services expenditure *not* consumed by EBPs to experiment and find out.

A final organizational source of resistance to EBPs is the view that they are too expensive to deliver, particularly at scale. This point is made by budget-holders at all levels of seniority, from national politicians to local service managers and whether providers or commissioners. It is particularly pertinent in the current climate of public sector austerity. Our response is that while EBPs are rarely a cheap option (although many are less expensive than expected), their cost should be weighed against their

potential financial benefits. These have been estimated for many programs using an economic model developed in the US and now being translated for the UK (Aos et al., 2011; Lee et al., 2012). For example, *Functional Family Therapy* has a cost-benefit ratio of 1:10 overall and yields significant benefits for the youth justice system in terms of lower crime costs (SRU, 2012). Challenges remain in terms of tracking and recovering such benefits, allocating them to the relevant agencies and disinvesting from some services initially to free up money for new programs, but the point is that thinking like an investor rather than a grant-giver changes the perception of cost.

6. Professional critiques

Professionals and their advocates offer a strong set of objections to EBPs – that such programs devalue the importance of relationships with clients, limit practitioners' autonomy, discourage innovation and, in some cases, are simply unsuitable or non-existent for the need in question. These are now considered in turn.

Foremost is the notion that relationships and reflective practice are critical to effective working with families and that the technical or instrumental nature of programs overrides these core practice elements (Jordan, 2008). Sometimes the argument is made even more forcefully, namely that the choice of intervention does not matter as long as there is a good therapeutic relationship (Rubin, 2011). Meta-analyses are selected to show that the quality of the practitioner-client relationship has a greater impact than the choice of intervention on client outcomes (Graybeal, 2007).

In response, we would point out that most programs require much face-to-face contact with users (Forrester, 2010) and that many emphasize in their manuals that a strong

therapeutic alliance is essential for program effectiveness (Rubin, 2011). But if relationships are important they are not *all important*. Effective interventions will generally work better if they are delivered with warmth, sensitivity and respect, but by itself relational competence or 'bedside manner' is unlikely to meet complex needs (Thompson, 2008). Meta-analyses suggesting that relationships are all that matters have methodological flaws: crucially, they include ineffective and even harmful interventions, cancelling out the effect of the effective interventions (Rubin, 2011). Moreover, other meta-analyses reach the opposite conclusion (Craighead et al., 2005), while even meta-analyses used to advocate for the centrality of relational support show that choice of intervention does matter, just not as much (Rubin, 2011).

A closely related critique is that EBPs 'de-professionalize' practitioners because they are overly didactic – they require implementing manuals and guidelines and deny professionals opportunities to improvise and tailor (Ayre & Calder, 2010). This is usually described as implementation with fidelity – delivering the program as intended. Critics maintain that it runs counter to what most people came into their professions to do and what they enjoy, which includes having discretion and the freedom to improvise (see Masson et al., 2008). It emasculates professional autonomy and discretion in the pursuit of standardising practice. This causes staff to focus more on compliance with procedures than on what they believe they should do to achieve positive outcomes for service users. Practitioners should be empowered to have individualized and responsive interactions with clients rather than being expected to complete a list of tasks (Chard & Ayre, 2010). There is a perceived risk of producing technicians who can follow directions to implement manualized interventions but who struggle to apply theoretical knowledge to complex practice situations (Adams et al.,

2009).

In response, we start from the premise that EBPs are designed based on evidence and rigorous testing (Fraser & Galinsky, 2010), and the results come from implementing them as intended: fidelity is important (Fixsen et al., 2005). However, programs exhibit different approaches to implementation. Some have scripted sessions or lesson plans that need to be followed sequentially whereas others are more flexible: for example, no course of *Functional Family Therapy* will be exactly the same because the intervention depends on the nature of the presenting problem. Program materials may be seen as supports not routines, and alterations are generally permitted and even encouraged, particularly as practitioners become more familiar with the program in question. For instance, the *PATHS* manual lists 'licensed alterations', urges the use of 'teachable moments' and incorporates many suggestions from teachers. Discretion and practitioner expertise, including theoretical knowledge, are still much needed – for example, to determine whether an EBP is suitable and for whom – and indeed are critical to the idea of evidence-based practice (Rubin, 2011). Further, even critics acknowledge the need for some expectations regarding quality and outcomes, so the issue is less about 'if' than 'how much' (Forrester, 2010).

Another rejoinder to the de-professionalization critique is that it takes great skill to deliver EBPs well. Put another way, program outcomes cannot surpass the skill level of implementers delivering them: *competence* of delivery is as important as *adherence* to content (Cross & West, 2011). For example, studies of the *Incredible Years* BASIC program have found that therapist skill contributed significantly to outcomes even after controlling for other aspects of the intervention, such as therapeutic alliance and

percentage of sessions attended (Scott et al., 2008; Eames et al., 2010). The program worked, but it worked better when delivered with more skill. There is also a difference between *being* highly skilled – which enables improvisation – and *becoming* so – which demands discipline to master skills (Forrester, 2010). This applies as much to work with children as it does to, say, cooking or sport. EBPs that achieve impressive results invest heavily in ensuring that practitioners are highly skilled or become so with training and supervision. Many practitioners are motivated by professional opportunities, intellectual challenge and the opportunity to master new skills (Stalker et al., 2007; Graham & Shier, 2010), so this should be a good fit. All of this may contribute to high reported levels of satisfaction among practitioners who have implemented well-known EBPs (e.g. Barnes et al., 2008).

Returning to the critiques, a further anxiety amongst practitioners is that focusing on established EBPs discourages innovation and creativity. Put another way, replicating established programs leaves little room for new ideas. This is lamented because, it is argued, there are many innovative programs that work but are not (yet) labelled 'evidence-based'. Staff in small non-governmental agencies worry that 'big' EBPs will push out the 'small' home-grown innovations they cannot afford to evaluate experimentally. The requirement for RCTs should, it is suggested, be relaxed. There is also a sense that clearinghouses of EBPs are elitist in that new programs struggle to break into a club of the usual suspects.

We accept the need for greater mobility in terms of new programs reaching high standards of evidence, but we do not accept that EBPs stifle innovation. Rather, EBPs are the *product* of innovation – often, incidentally, by practitioners (the developers of

most of the best known programs are or were practitioners). EBPs are the result of an intensive process of design and testing: this is how innovation works (Bessant & Tidd, 2009). Moreover, innovation is ongoing, with many accommodations or adaptations continuing to be made for different groups and contexts or in response to findings (eg. Webster-Stratton & Reid, 2010).

Further, innovations are not intrinsically good. In industry, for every idea that results in a robust, marketable product, many will have been discarded. The same should apply in children's services. To illustrate, the majority of programs submitted for possible funding to a recent UK funding initiative lacked sound logic models and were deemed unlikely to work (Hobbs, 2012). Many were seeking to do badly what several EBPs already do well, so energy is arguably better focused elsewhere.

Indeed, there are neglected areas of children's services crying out for creativity. For example, there are relatively few EBPs proven to prevent or treat obesity, abuse and neglect, or suicidal ideation, and hardly any that promote entry into stable post-secondary education, training or employment. But innovation can also occur in how EBPs are delivered. For instance, practitioners invariably struggle to engage parents in evidence-based parenting programs (Axford et al., 2012b). And even good innovation is more than having smart ideas: it is an incremental process of theorising, design and testing to find out if something is needed, effective, liked, wanted, used and scalable (Fraser et al., 2009; Little, 2012). The history of innovation shows that fields advance when new ideas build on, extend and only occasionally depart from what exists (Johnson, 2010).

A final professional critique is that there may not be a suitable EBP for the need or group in question. Perhaps the problem is too complex, such as serious or prolonged maltreatment, or the needs are too heterogeneous or co-exist with other problems (Thoburn, 2010). A variation of this objection is practitioners saying that 'my children/parents are unique'; in other words, the program has not been tested with the users in question, so it is not proven for them. Another variation is that children and families – particularly those in greatest need – do not like and will not use programs because they are too structured, didactic, demanding and lengthy.

We would offer several responses. It is true that EBPs do not exist for some needs. When programs are mapped against key developmental outcomes affected there are significant gaps, as noted already. These need to be addressed through innovation and testing. The corollary of this is that EBPs *do* target and have an impact on many developmental outcomes. They are also being tailored for and tested with different populations, such as children with special educational needs or severe behavioral or emotional problems as well as those in the child welfare system or from different cultures (eg. Webster-Stratton & Reid, 2010). Evaluations increasingly examine not just whether a program works but also whom it works for and under what conditions, with a view to potentially making modifications (Chauveron & Perkins, 2009).

It should also be noted that if a suitable program does not exist it is unhelpful to impose one that it is unsuitable. Indeed, a central tenet of the evidence-based practice approach is that conclusive evidence of effectiveness may not exist and the choice of intervention might need to be based on the best evidence available, a practitioner's theoretical knowledge or their experience of what works with their population. This

reinforces the earlier point about EBPs demanding and extending professionalism.

Finally, while some users may dislike some programs, participants in EBPs are often very positive about both the process and its impact (eg. Barnes et al., 2008). Some caution is required because user-satisfaction surveys tend to reflect the views of users who have stayed the course, which is likely to lead to bias. But the alleged resistance from users is often exaggerated.

7. Conclusion

In setting out critiques of EBPs we are conscious of the danger of appearing overly negative. We have sought to indicate that by no means all of the objections are fair and that efforts are being made to address the others. The best EBPs, when delivered well, make life considerably better for the children and families who receive them and the people with whom they interact. Often children with the most serious needs make the largest gains. Society is made a little fairer and a little less violent as a result, and public money is used more efficiently. We reiterate our earlier point: EBPs are underutilized and should be used much more. We end this article by discussing implications of our appraisal, starting with evaluation and program design.

Where appropriate, and feasible, programs should be subjected to good RCTs. Otherwise, good QED studies should be used. More such evaluations need to be in the real world. In turn, there is a need to develop alternative methods for robust impact evaluation (eg. Ford et al., 2009). For programs already proven to work in one context replication studies are needed to see if the outcomes can be replicated (Little et al., 2012). Evaluators should take care not to exaggerate impact, particularly when

outcomes are patchy. The reporting of studies should be more transparent so that readers can judge their quality, and potential conflict of interest from developer involvement should be acknowledged. The standard for what constitutes 'evidence-based' should rise to the point where there is an international standard for what constitutes proven and replicable.

Methods for moving from an innovation to a proven model need to be developed and tested. There is also a need to distinguish better through empirical studies the difference between core and adaptable elements in programs, so that there can be an emphasis on implementing the core well while innovating with the rest. Experiments that compare different versions of programs are vital. Further, since programs are not a panacea, evidence-based policies, processes and practices need to be identified. This will mean, for example, testing 'kernels' prospectively.

Our appraisal of critiques of EBPs also has implications for the fit between programs and service systems. Researchers need to learn more about what makes programs transportable and understand cultural variation in impact. Methods for adapting EBPs need to be developed and tested. When introducing programs to systems it is imperative to stress the importance of the practitioner, demonstrating how professional knowledge and experience contribute to outcomes. The place for innovation should be highlighted, as should the radical nature of programs that reach the most needy and reduce inequality, and the significant amount of 'face time' with users. Program training should be seen as an opportunity to master new skills and certified as such. These are all things that practitioners are reported to find motivating. Systems also need incentives. There is a need to show, for example, how improved

outcomes – defined in terms of child health and development – can change outputs – indicators of how children are processed in a system. This is what systems care about. Rather than 'push' programs to systems, the aim should be to create 'pull' and make systems want programs. In addition to making programs 'system ready', systems need to become 'program ready'. For example, more investment is needed in staff having resources and skills to implement EBPs. Innovations in contracting, like payment by results and social impact bonds, are likely to make systems more outcomes-focused and therefore a better a fit for EBPs.

Lastly, our analysis informs some broader recommendations. First, we reaffirm the ethical principle of 'first, do no harm'. Too often, the assumption in our field is that good intentions lead to better outcomes. This is not always the case. In our field we need to be more skeptical about our ability to make things better, to have lower (more realistic) expectations of impact and, crucially, to stop harmful interventions. Second, providers need to become more accountable for outcomes and costs. Most EBPs pass the 'my child test' – would I be happy for my child to receive that program? Many also pass the 'my money test' – would I invest in it? Such programs are underutilized, however, while many that would fail such tests are widely used. Third, there is a need to identify and, where appropriate, adopt EBPs regardless of their country of origin. Many currently come from the US but the development of programs in Europe and elsewhere that meet high standards of evidence should be encouraged. Fourth, the delivery of EBPs in regular services need not be accompanied by onerous data collection. Rather, data for monitoring outcomes and implementation should be collected sparingly – that is, where necessary and useful. Finally, intermediaries, including clearinghouses of EBPs, play an important role as the honest broker by

providing commissioners and policy makers with the information they need to make decisions.

Acknowledgements

We are very grateful to colleagues who kindly took the time to provide thorough and helpful comments on an earlier draft of this article: Roger Bullock, Tracey Bywater, Bette Chambers, Donald Forrester, and Andrew Morris.

References

Adams, K. B., Matto, H. C., & Lecroy, C. W. (2009). Limitations of evidence-based practice for social work education: Unpacking the complexity. *Journal of Social Work Education* 45(2), 165–186.

Aos, S., Lee, S., Drake, E., Pennucci, A., Klima, T., Miller, M., Anderson, L., Mayfield, J., & Burley, M. (2011). *Return on investment: Evidence-based options to improve statewide outcomes. Document No. 11-07-1201*. Olympia: Washington State Institute for Public Policy.

Atria, M., & Spiel, C. (2007). The Viennese Social Competence (ViSC) training for students: Program and evaluation. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization and peer harassment: A handbook of prevention and intervention*. (pp. 179–198). New York, NY: Haworth Press.

Axford, N. (2008). *Exploring concepts of child wellbeing: Implications for children's services*. Bristol: Policy Press.

Axford, N., Elliott, D. S. and Little, M. (2012a). Blueprints for Europe: promoting evidence-based programmes in children's services. *Psychosocial Intervention* 21 (2), 205-214.

Axford, N., Lehtonen, M., Tobin, K., Kaoukji, D. & Berry, V. (2012b). Engaging parents in parenting programs: lessons from research and practice. *Children and Youth Services Review*. 34 (10), 2061-2071.

Ayre, P., & Calder, M. C. (2010). The de-professionalisation of child protection: regaining our bearings. In: P. Ayre & M. Preston-Shoot (Eds.), *Children's Services at the Crossroads: A Critical Evaluation of Contemporary Policy for Practice*. Lyme Regis, UK: Russell House Publishing.

Barnes, J., Ball, M., Meadows, P., McLeish, J., Belsky, J. & the FNP Implementation Research Team (2008) *Nurse-Family Partnership Programme: First Year Pilot Sites. Implementation in England Pregnancy and the Post-partum Period*. Research Report DCSF-RW051. DCSF & DoH.

Barth, R. P., Lee, B, R., Lindsey, M. A., Collins, K. S., Strieder, F., Chorpita, B. F., Becker, K., & Sparks, J. A. (2012). Evidence-based practice at a crossroads: the emergence of common elements and factors. *Research on Social Work Practice, 22*(1), 108-119.

Bessant, J. & Tidd, J. (2007) *Innovation and Entrepreneurship*, Chichester, Wiley.

Bickman, L., Douglas Kelley, S., Breda, C., Regina de Andrade, A. & Riemer, M. (2011) 'Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial', *Psychiatric Services* 62 (12), 1423-1429.

Botvin, G. J., & Griffin, K. W. (2005). Prevention science, drug abuse prevention, and Life Skills Training: Comments on the state of the science. *Journal of Experimental Criminology, 1*, 63-78.

Bradshaw, J. (Ed) (2011). *The Well-being of Children in the UK*. Bristol, UK: Policy Press.

Bullock, R., Little, M. and Millham, S. (1992) 'The Relationships between Quantitative and Qualitative Approaches to Social Policy Research', in Brannen, J. (ed.) *Mixing Methods: qualitative and quantitative research*, Aldershot, Avebury.

Bumbarger, B. K., & Perkins, D. F. (2008). After randomised trials: issues related to the dissemination of evidence-based interventions. *Journal of Children's Services,* 3 (2), 55-64.

Butler, S., Baruch, G., Hickey, N & Fonagy, P. (2011). A Randomized Controlled Trial of Multisystemic Therapy and a Statutory Therapeutic Intervention for Young Offenders. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*(12), 1220-1235.

Case, S. (2006). Young people 'at risk' of what? Risk-focused early intervention as crime prevention. *Youth Justice, 6*, 171-178.

Case, S. (2007). Questioning the 'evidence' of risk that underpins evidence-led youth justice interventions. *Youth Justice, 7*, 91-105.

Cash, S. J., Ingram, S. D., Biben, D. S., McKeever ,S. J., Thompson, R. W. & Ferrell, J. Z. (2012) 'Moving forward without looking back: performance management systems as real-time evidence-based practice tools', *Children and Youth Services* Review 34, 655-659.

Catalano, R. F., & Hawkins, J. D. (1996). The social development model: A theory of antisocial behavior. In J. D. Hawkins (Ed.), *Delinquency and Crime: Current Theories* (pp. 149-197). New York, NY: Cambridge University Press.

Chard, A., & Ayre, P. (2010). Managerialism: At the tipping point? In: P. Ayre & M. Preston-Shoot (Eds.), *Children's Services at the Crossroads: A Critical Evaluation of Contemporary Policy for Practice.* Lyme Regis, UK: Russell House Publishing.

Chauveron, L., & Perkins, D. (2009). Prevention in context: classroom heterogeneity as a moderator of PATHS' effectiveness. *Journal of Children's Services, 4*(1), 44-59.

Chen, X., & Eisenberg, N. (2012). Understanding cultural issues in child development: Introduction. *Child Development Perspectives, 6*(1), 1-4.

Collishaw, S. (2012) 'Time trends in young people's emotional and behavioural problems', in Hagell. A. (Ed.) *Changing Adolescence: Social Trends and Mental Health*. Bristol: Policy Press.

Craighead, W. E., Sheets, E. S., & Bjornsson, A. S. (2005). Specificity and nonspecificity in psychotherapy. *Clinical Psychology: Science and Practice*, 12(2), 189–193.

Cross, W., & West, J. (2011). Examining implementer fidelity: conceptualising and measuring adherence and competence. *Journal of Children's Services, 6*(1), 18-33.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D. & Schellinger, K. B. 2011. The impact of enhancing student's social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development* 82 (1), 405-432.

Eames, C., Daley, D. Hutchings, J., Whitaker C.J., Hughes, C., Jones, K., Hughes, K., & Bywater, T. (2010). The impact of group leaders behavior on parents acquisition of key parenting skills during parent training. *Behavior Research and Therapy*, 48 (12), 1221-1226 doi:10.1016/j.brat.2010.07.011

Ecclestone, K., & Hayes, D. (2009). *The dangerous rise of therapeutic education*. Abingdon, UK: Routledge

Eisner, M. (2009a). No effects in independent prevention trials: can we reject the cynics' hypothesis? *Journal of Experimental Criminology, 5*, 163-183.

Eisner, M. (2009b). Reply to the comments by David Olds and Lawrence Sherman. *Journal of Experimental Criminology, 5*, 215-218.

Eisner, M. & Nivette, A. (2012). How to reduce the global homicide rate to 2 per 100,000 by 2060. In Welsh, B. & Loeber, R. (Eds). *The Future of Criminology*. Oxford: Oxford University Press.

Eisner, M., Nagin, D., Ribeaud, D. & Malti, T. (2012). Effects of a universal parenting program for highly adherent parents: a propensity score matching approach. *Prevention Science*, Published online 12th January DOI 10.1007/s11121-011-0266-x.

Elliott, D. (2010). Lessons from Columbine: effective school-based violence prevention strategies and programmes. *Journal of Children's Services, 4*(4), 53-62.

Embry, D. D., & Biglan, A. (2008). Evidence-based Kernels: Fundamental units of behavioral influence. *Clinical, Child and Family Psychology Review, 11*(3), 75-113.

Faggiano, F., Galanti, M. R., Bohrn, K., Burkhart, G., Vigna-Taglianti, F., Cuomo, L., Fabiani, L., Panella, M., Perez, T., Siliquini, R., van der Kreefts, P., Vassara, M., Wiborg, G., EU-Dap Study Group (2008). The effectiveness of a school-based substance abuse prevention programme: EU-Dap cluster randomised controlled trial. *Preventive Medicine* 47 (5), 537-543.

Farrington, D. P. (2003). A short history of randomized experiments in criminology: a

meager feast. *Evaluation Review* 27 (3), 218-227.

Ferrer-Wreder L, Stattin H, Lorente CC, Tubman JG & Adamson L (2004) *Successful prevention and youth development programs. Across borders*. New York: Kluwer Academic/ Plenum Publishers.

Fixsen, D., Naoom, S. F., Blasé, K. A., Friedman, R. M. & Wallace, F. 2005. *Implementation Research: A Synthesis of the Literature*. Tampa, USF.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Moscicki, E. K., Schinke, S., Valentine, J. C., & Peter, J. I. (2005). Standards of Evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*(3), 151-175.

Ford, T., Hutchings, J., Bywater, T., Goodman, A. & Goodman, R. (2009) Strengths and Difficulties Questionnaire Added Value Scores: evaluating effectiveness in child mental health interventions, *British Journal of Psychiatry 194,* 552-558.

Forrester, D. (2010). Playing with fire or rediscovering fire? The perils and potential for evidence based practice in child and family social work. In: P. Ayre & M. Preston-Shoot (Eds.), *Children's Services at the Crossroads: A Critical Evaluation of Contemporary Policy for Practice.* Lyme Regis, UK: Russell House Publishing.

Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: designing and developing social programs. *Research on Social Work Practice, 20*(5), 459-466.

Furlong, McGilloway S, Bywater T, Hutchings J, Smith SM, Donnelly M (2012) *Behavioral and Cognitive-behavioral Group-based Parenting Programmes for Early-onset Conduct Problems in Children Aged 3 to 12 Years (Review)*. Cochrane Collaboration, Wiley.

Gallart, S. C., & Matthey, S. (2005). The effectiveness of Group Triple P and the impact of the four telephone contacts. *Behaviour Change, 22*(2), 71-80.

Garrett, P. M. (2009). *Transforming children's services? Social work, neoliberalism and the 'modern' world*. Maidenhead: Open University Press/ McGraw-Hill.

Gill, C. E. (2011). Missing links: how descriptive validity impacts the policy relevance of randomized controlled trials in criminology. *Journal of Experimental Criminology, 7,* 201-224.

Gorman, D. M. (2002a). The "science" of drug and alcohol prevention: the case of the randomized trial of the Life Skills Training Program. *International Journal of Drug Policy, 13*, 21-26.

Gorman, D. M. (2002b). Defining and operationalizing 'research-based' prevention: a critique (with case studies) of the US Department of Education's safe, disciplined and drug-free schools exemplary programs. *Evaluation and Program Planning, 25*, 295-302.

Gorman, D. M. (2003). The best of practices, the worst of practices: The making of science-based primary prevention programs. *Psychiatric Services, 54*(8) 1087-1089.

Gorman, D. M. (2005a). Drug and violence prevention: Rediscovering the critical rational dimension of evaluation research, *Journal of Experimental Criminology, 1*, 39-62.

Gorman, D. M. (2005b). The centrality of critical rational reasoning in science: A response and further comments on the Life Skills Training program, the Seattle Social Development Project, and prevention studies. *Journal of Experimental Criminology, 1*, 263-275.

Graham, J. R. & Shier, M. L. (2010) 'The social work profession and subjective well-being: the impact of a profession on overall subjective well-being', *British Journal of Social Work* 40 (5), 1553-1572.

Gray, M., Plath, D., & Webb, S. A. (2009). Evidence-based social work: A critical stance. Abingdon: Routledge

Graybeal, C. T. (2007). Evidence for the art of social work. *Families in Society*, 88(4), 513–523.

Grietens, H. (2010). Discerning European perspectives on evidence-based interventions for vulnerable children and their families. *International Journal of Child and Family Welfare, 13*(1–2), 6–17.

Haggerty, K. P., Kosterman, R., Catalano, R. F., & Hawkins, J. D. (1999, July). Preparing for the Drug Free Years. *Office of Juvenile Justice and Delinquency Prevention Juvenile Justice Bulletin [NCJ 173408]*. Retrieved from https://www.ncjrs.gov/pdffiles1/173408.pdf

Hahlweg, K., Heinrichs, N., Kuschel, A., Bertram, H., & Nauman, S. (2010). Long-term outcome of a randomized controlled universal prevention trial through a positive parenting program: Is it worth it? *Child and Adolescent Psychiatry and Mental Health*, 4, 1-14.

Harvey, D. (2005). *A brief history of neoliberalism*. Oxford: Oxford University Press.

Hawkins, J. D., & Catalano, R. F. (2005). Doing prevention science: A response to Dennis M. Gorman and a brief history of the quasi-experimental study nested within the Seattle Social Development Project. *Journal of Experimental Criminology, 1*, 79-86.

Henggeler, S. W., & Schoenwald, S. K. (2011). Evidence-based interventions for juvenile offenders and juvenile justice policies that support them. *SRCD Social Policy Report, 25*(1), 3-20.

Hobbs, T. (2012) *Realising Ambition: Applying standards of evidence to preventive interventions for anti-social and offending behavior*, Paper presentation at Society for Prevention Research Annual Meeting, Washington, US, 31st May.

Holder, H. (2009). Prevention programs in the 21st century: what we do not discuss in public. *Addiction, 105*, 578-581.

Hutchings, J., Bywater, T., Daley, D., Gardner, F., Jones, K., Eames, C., Tudor-Edwards, R. & Whitaker, C. (2007). A pragmatic randomised control trial of a parenting intervention in sure start services for children at risk of developing conduct disorder. *British Medical Journal, 334, 678 - 682.*

Ipsos MORI (2011) *Children's Well-being in UK, Sweden and Spain: the Role of Inequality and Materialism*. London: Ipsos MORI.

Johnson, S. (2010) *Where Good Ideas Come From: The Natural History of Innovation*. New York: Penguin.

Jordan, B. (2008). Social value in policies for children: contract or culture? *Journal of Children's Services 3(3)*, 65-75.

Kratochwill, T. R., McDonald, L., Levin, J. R., Young Bear-Tibbets, H., & Demaray, M. K. (2004). Families and Schools Together: an experimental analysis of a parent-mediated multi family group program for American Indian Children. *Journal of School Psychology, 42*, 359-383.

Kumpfer, K. L., Magalhães, C., & Xie, J. (2012). Cultural adaptations of evidence-based family interventions to strengthen families and improve children's

developmental outcomes. *European Journal of Developmental Psychology, 9*(1), 104-116

Larsson. B., Fossum, B., Clifford, G., Drugli, M., Handegard, B., Morch. W. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial', *European Child & Adolescent Psychiatry*, 18(1), 42-.-52.

Lee, S., Drake, E., Pennucci, A., Bjornstad, G., & Edovald, T. (2012). Economic evaluation of early childhood education in a policy context. *Journal of Children's Services, 7*(1), 53-63.

Lipsey, M., Howell, J. C., Kelly, M. R., Chapman, G., Carver, D. (2010*). Improving the effectiveness of juvenile justice programs: A new perspective on evidence-based practice*. Center for Juvenile Justice Reform: Georgetown Public Policy Institute, Georgetown University.

Little, M. (2010). *Proof Positive*. London: Demos

Little, M. (2011) *How to Achieve Lasting Impact at Scale: Behavior Change and the Spread of Family Health Innovations in Low-Income Countries*. Synthesis and summary of a convening hosted by the Bill & Melinda Gates Foundation in Seattle, November 1-2. Dartington: Social Research Unit.

Little, M. (2012) *From Innovation to Proven Model*. Dartington: Social Research Unit.

Little, M., Axford, N., & Morpeth, L. (2004). Risk and protection in the context of services for children in need. *Child and Family Social Work, 9*(1), 105-118.

Little, M., Axford, N., Morpeth, L., & Weyts, A. (2005). Evaluating children's services: recent conceptual and methodological developments. *British Journal of Social Work, 35*(1), 73-88.

Little, M., & Maughan, B. (Eds.). (2010). *Effective interventions for children in need*. Aldershot: Ashgate.

Little, M., Berry, V., Morpeth, L., Blower, S., Axford, N., Taylor, R., Bywater, T., Lehtonen, M. & Tobin, K. (2012) The impact of three evidence-based programmes delivered in public systems in Birmingham, UK. Article submitted to *International Journal of Conflict and Violence.*

Little, M., & Axford, N. (2012). Taking over the world or scratching at its surface? *Journal of Children's Services 6*(4), 207-209.

Malti, T., Ribeaud, D. & Eisner, M. (2011). The Effectiveness of Two Universal Preventive Interventions in Reducing Children's Externalizing Behavior: A Cluster Randomized Controlled Trial. *Journal of Clinical Child & Adolescent Psychology 40*(5), 677–692**.**

Masson, H., Frost, N. & Parton, N. (2009). Reflections from the 'front-line': Social workers' experiences of post qualifying child care training and their current work practices in the new children's services, *Journal of Children's Services* 3 (3), 54-64.

Maughan B, Collishaw S, Meltzer H and Goodman R. 2008. Recent trends in UK child and adolescent mental health. *Social Psychiatry and Psychiatric Epidemiology* 43, 305-310.

Maughan, B., & Little, M. (Eds.) (2010). *Child Development*. Aldershot: Ashgate.

McGilloway S, Mhaille G, Bywater T, Furlong M, Leckey Y, Kelly P, Comiskey C, and Donnelly M (2012), 'A parenting intervention for childhood behavioural problems: a randomized controlled trial in disadvantaged community-based settings', *Journal of Consulting and Clinical Psychology*, Vol 80(1), 116-127.

Milova, H. J-L. & Sohere, S. (2011) 'Key concepts of parenting support in France and Germany', *International Journal of Child & Family Welfare* 14 (3-4), 105-118.

Misurell, J. R., & Springer, C. (2011). Developing Culturally Responsive Evidence-Based Practice: A Game-Based Group Therapy Program for Child Sexual Abuse (CSA). *Journal of Child and Family Studies*, DOI: 10.1007/s10826-011-9560-2

Mitchell, P. F. (2011) 'Evidence-based practice in real-world services for young people with complex needs: new opportunities suggested by recent implementation

science', *Children and Youth Services Review* 33, 207-216.

NICE (National Institute for Health and Clinical Excellence) (2006) *Parent Training/Education Programmes in the Management of Children with Conduct Disorders*. London: NICE.

Ogden, T. and Hagen, K. A. (2008). Treatment effectiveness of Parent Management Training in Norway: A randomized controlled trial of children with conduct problems. *Journal of Consulting and Clinical Psychology*, 76(4), 607-621.

Olds, D. (2009). In support of disciplined passion. *Journal of Experimental Criminology, 5*, 201-214.

Olsson, T. (2010) MST with conduct disordered youth in Sweden: costs and benefits after 2 years. *Research on Social Work Practice, 20*(6), 561-571.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation.* London: Sage.

Penn, H. (2002). The World Bank's view of early childhood. *Childhood, 9*(1), 118-132.

Perry, A. E., Weisburd, D., & Hewitt, C. (2010). Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials, *Journal of Experimental Criminology 6*, 245-262.

Petrie, P., Boddy, J., Cameron, C., Wigfall, V. & Simon, A. (2006). *Working with Children in Care – European Perspectives.* Maidenhead: Open University Press.

Petrosino, A., Turpin-Petrosino, C. & Buehler, J. (2003) Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: a systematic review of the randomized experimental evidence, *Annals of the American Academy of Political and Social Science 589(1)*, 41-62.

Ritter, G.W. (2012). Using the proper tool for the task: RCTs *are* the gold standard for estimating programme effects – a response to Stewart-Brown et al. *Journal of Children's Services, 7*(2).

Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach 6th edition.* Thousand Oaks, CA: Sage.

Rowlands, J. (2010). Services are not enough: child well-being in a very unequal society. *Journal of Children's Services, 5*(3), 80-88.

Rubin, A. (2011). Teaching EBP in social work: retrospective and prospective. *Journal of Social Work, 11*(1), 64-79.

Salmivalli, C., Kärnä, A., and Poskiparta, E. (2011). Counteracting bullying in Finland: The KiVa program and its effects on different forms of being bullied. International Journal of Behavioral Development, doi:10.1177/0165025411407457

Scott, S., Carby, A. & Rendu, A. (2008) *Impact of Therapists' Skill on Effectiveness of Paretning Groups for Child Antisocial Behavior*. Unpublished paper.

Scott, S., O'Connor, T. G., Futh, A., Matias, C., Price, J., Doolan, M. (2010). Impact of a parenting program in a high-risk, multi-ethnic community: the PALS trial. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 51*(12), 1331-1341.

Shadish, W. R. (2011) Randomized controlled studies and alternative designs in outcome studies: challenges and opportunities. *Research on Social Work Practice 21* (6), 636-643.

Sheldon, B. & Macdonald, G. (2009). *A Textbook of Social Work*. London: Routledge.

Sherman, L. W. & Strang, H. (2009). Testing for analysts' bias in crime prevention experiments: can we accept Eisner's one-tailed test? *Journal of Experimental Criminology 5*, 185-200.

Soydan, H., Mullen, E. J., Alexandra, L., Rehnman, J., & Li, Y-P (2010). Evidence-based clearinghouses in social work. *Research on Social Work Practice, 20*(6), 690-700.

Spiel, C., & Strohmeier, D. (2012). Evidence-based practice and policy: When researchers, policy makers, and practitioners learn how to work together. *European Journal of Developmental Psychology, 9*(1), 150-162.

SRU (Social Research Unit) (2012). *Investing in Children: Youth Justice 1.1*. Dartington: SRU.

Stalker, C.A., Mandell, D., Frensch, K., Harvey, C. & Wright, M. (2007). Child welfare workers who are exhausted yet satisfied with their jobs: How to do they do it? *Child & Family Social Work,* 12 (2), 182-191.

Stemmler, M., Kötter, C., Bühler, A., Jaursch, S., Beelmann, A. & Lösel, F. (2013). Prevention of familial transmission of depression through a family-oriented programme targeting parenting as well as the child's social competence. *Journal of Children's Services* 8 (1).

Sundell, K., Hansson, K., Lofholm, C. A., Olsson, T., Gustle, L. H., & Kadesjo, C. (2008). The Transportability of Multisystemic Therapy to Sweden: Short-Term Results from a Randomized Trial of Conduct-Disordered Youths. *Journal of Family Psychology*, *22*(3), 550-560.

Stewart-Brown, S., Anthony, R., Wilson, L., Winstanley, S., Stallard, N., Snooks, H., & Simkiss, D. (2011). Should randomised controlled trials be the ''gold standard'' for research on preventive interventions for children? *Journal of Children's Services,* 6(4), 228-235.

Thoburn, J. (2010). Towards knowledge-based practice in complex child protection cases: A research-based expert briefing. *Journal of Children's Services, 5*(1), 9-24.

Thompson, N. (2008). Focusing on outcomes: Developing systematic practice. *Practice* 20(1), 5-16.

Tolan, P., Gorman-Smith, D & Henry, D. (2004) 'Supporting families in a high-risk setting: proximal effects of the SAFEChildren preventive intervention', *Journal of Consulting and Clinical Psychology* 72, 855-869.

Torgerson, D., & Torgerson, C. (2008). Randomised controlled trials in children's services. *Journal of Children's Services, 3*(1), 2-8.

UNICEF (2007). *Child poverty in perspective: An overview of child wellbeing in rich countries*. Florence: UNICEF Innocenti Research Centre.

Webb, S.A. (2001) 'Considerations on the validity of evidence-based practice in social Work' *British Journal of Social Work* 31 (1), 57-79.

Webster-Stratton, C., Kolpacoff, M., & Hollinsworth, T. (1988). Self-administered videotape therapy for families with conduct-problem children: Comparison with two cost-effective treatments and a control group. *Journal of Consulting Clinical Psychology*, 56, 558-566.

Webster-Stratton, C. & Reid, M. J. (2010). Adapting The Incredible Years, an evidence-based parenting programme, for families involved in the child welfare system. *Journal of Children's Services 5*(1), 25-42.

Wilkinson, R. & Pickett, K. (2009) *The Spirit Level: Why more equal societies almost always do better*. London: Allan Lane.